UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 14-25
Statistics and Econometrics Series (17)
September 2014

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

# A Bootstrap Likelihood approach to Bayesian Computation

Weixuan Zhu [1], J.Miguel Marín[1], Fabrizio Leisen[2]

**Abstract**

Recently, an increasingly amount of literature focused on Bayesian computational methods to address problems with intractable likelihood. These algorithms are known as *Approximate Bayesian Computational* (ABC) methods. One of the problems of these algorithms is that the performance depends on the tuning of some parameters, such as the summary statistics, distance and tolerance level.

To bypass this problem, an alternative method based on empirical likelihood was introduced by Mengersen et al. (2013), which can be easily implemented when a set of constraints, related with the moments of the distribution, is known.

However, the choice of the constraints is crucial and sometimes challenging in the sense that it determines the convergence property of the empirical likelihood. To overcome this problem, we propose an alternative method based on a bootstrap likelihood approach. The method is easy to implement and in some cases it is faster than the other approaches. The performance of the algorithm is illustrated with examples in Population Genetics, Time Series and a recent non-explicit bivariate Beta distribution. Finally, we test the method on simulated and real data random fields.

***Keywords***: Approximate Bayesian Computational methods, bootstrap likelihood, empirical likelihood, bivariate Beta distribution, population genetics.

1 Departamento de Estadística, Universidad Carlos III de Madrid,
e-mail: weixuan.zhu@uc3m.es
jmmarin@est-econ.uc3m.es.
2 School of Mathematics, Statistics and Actuarial Sciences, University of Kent, UK.
e-mail: fabrizio.leisen@gmail.com

# A Bootstrap Likelihood approach to Bayesian Computation

Weixuan Zhu[1], J. Miguel Marín[1], Fabrizio Leisen[1,]

[a]*Departamento de Estadistica, Universidad Carlos III de Madrid, Spain*
[b]*School of Mathematics, Statistics and Actuarial Sciences, University of Kent, UK*

**Abstract**

Recently, an increasingly amount of literature focused on Bayesian computational methods to address problems with intractable likelihood. These algorithms are known as *Approximate Bayesian Computational* (ABC) methods. One of the problems of these algorithms is that the performance depends on the tuning of some parameters, such as the summary statistics, distance and tolerance level. To bypass this problem, Mengersen, Pudlo and Robert (2013) introduced an alternative method based on empirical likelihood, which can be easily implemented when a set of constraints, related with the moments of the distribution, is known. However, the choice of the constraints is crucial and sometimes challenging in the sense that it determines the convergence property of the empirical likelihood. To overcome this problem, we propose an alternative method based on a bootstrap likelihood approach. The method is easy to implement and in some cases it is faster than the other approaches. The performance of the algorithm is illustrated with examples in Population Genetics, Time Series and a recent non-explicit bivariate Beta distribution, see Arnold and Ng (2011) and Crackel and Flegal (2014). Finally, we test the method on simulated and real data random fields.

*Keywords:* Approximate Bayesian Computational methods, bootstrap likelihood, empirical likelihood, bivariate Beta distribution, population genetics

## 1. Introduction

The last decade has seen a growing stream of literature to address Bayesian inference when the likelihood is intractable. With the word *intractable* we mean that the likelihood function is analytically infeasible or computationally intensive.

Such likelihoods naturally arise in many research areas, for instance in Population Genetics (Beaumont et al. (2002), Drovandi and Pettitt (2010)), Epidemics (McKinley, Cook and Deardon (2009)) and Hidden Markov Models (Dean et al. (2014)). The typical estimation approach is based on algorithms that are customary called *approximate Bayesian computational* (ABC) methods.

*Corresponding author at: School of Mathematics, Statistics and Actuarial Sciences Cornwallis Building, University of Kent, Canterbury, CT2 7NF, UK.

Despite of its popularity, these algorithms suffer calibration problems (like summary statistics, distance, tolerance) and recently some papers focused on algorithms that try to bypass some of the tuning problems, see for instance Fearnhead and Prangle (2012), Del Moral, Doucet and Jasra (2012) and Sisson et al. (2007).

Mengersen, Pudlo and Robert (2013) proposed an alternative approach that deals with the well-established empirical likelihood approximation. The authors apply the method in a Bayesian framework to avoid the choice of the ABC parameters and, with the acronym $BC_{el}$, they mean *Bayesian computation with empirical likelihood*. The basic algorithm works in the following way: Firstly, generate $M$ parameters $\theta_i$ from the prior distribution. Then, set the weight $\omega_i = L_{el}(\theta_i|y)$, where $L_{el}(\theta_i|y)$ is the empirical likelihood of $\theta_i$ given the observed data $y$. The output of $BC_{el}$ is a sample of size $M$ of parameters with associated weights, which operate as an importance sampling output. However, the validation of the empirical likelihood depends on the choice of a set of constrains that ensures convergence. Our work is in a similar flavour of Mengersen, Pudlo and Robert (2013). In particular, we propose to use bootstrap likelihood (Davison, Hinkley and Worton (1992)) instead of empirical likelihood since this method is an automatic procedure that is easy to implement and doesn't require to set any objective constrain. Furthermore, as the empirical likelihood, the bootstrap likelihood converges to the true likelihood as the number of observations increases.

The basic bootstrap likelihood method, introduced by Davison, Hinkley and Worton (1992), is summarized in section 2. Section 3 is devoted to the description of our methodology and in section 4 the methodology is tested on several examples such as Time Series, Population Genetics, and the bivariate Beta distribution of Arnold and Ng (2011). Finally, we test the method on simulated and real data random fields.

## 2. Bootstrap Likelihood

Davison, Hinkley and Worton (1992) use nested bootstrap calculation in conjunction with kernel smoothing methods to calculate estimates of the density of a given statistic for a range of parameter values. These density estimates are used to generate values of an analogue of a likelihood function by curve-fitting methods. Specifically, the basic method can be summarized as follows:

- Suppose $\theta$ is the parameter of interest and $\hat{\theta}$ is the parameter estimated by its sample analogue. Take $K$ bootstrap samples of size $n$ (same size as the original data) to obtain a series of populations $P_1^*, ..., P_K^*$ with parameter values $\hat{\theta}_1^*, ..., \hat{\theta}_K^*$ (first-level bootstrap). If the parameters do not have an explicit estimator, we estimate them by using numerical methods.

- For the $i$-th bootstrap sample we take $L$ samples from $P_i^*$, where $L$ is preferably 1000 (as suggested in Davison, Hinkley and Worton (1992)). For each sample calculate the analogue of $\theta$, denoted by $\hat{\theta}_{ij}^{**}$ (second-level bootstrap). Use a density kernel estimation to estimate the density of $\theta$ at points $\hat{\theta}_i^*$, say $g(\hat{\theta}_i^*)$, then the $K$ values $l(\theta_i^*) = \log[g(\hat{\theta}_i^*)]$ are obtained.

- Apply a smooth curve-fitting algorithm to obtain the whole log bootstrap likelihood curve.

For some dependent data, the outlined method also applies. For example, in regression-type problems it may be possible to write $y_i = \eta(x_i, \varepsilon_i)$, where the errors $\varepsilon_i$ are assumed to have a common distribution $F$. The same bootstrap procedure could be employed with respect to $\varepsilon_i$ and bootstrap samples are calculated as $y_i^* = \eta(x_i, \varepsilon_i^*)$ with $\varepsilon_i^*$ independently sampled from $\hat{F}$. This is a typical situation in dynamic models, see for instance, Example 4.1.

The relation between empirical likelihood and bootstrap likelihood is also explored from a theoretical point of view in Davison, Hinkley and Worton (1992). They show that in the case of an estimator determined by a monotonic estimating function, standardized so that the leading term is of order one, empirical likelihood and bootstrap likelihood agree to order $n^{-\frac{1}{2}}$ in the number of observations, $n$.

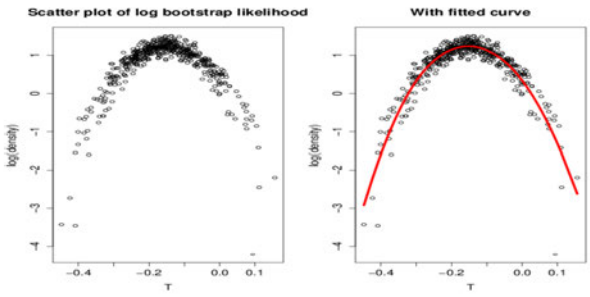


Figure 1: The left figure is plotted after the first two steps in the above summarized method. Basically, the first-level bootstrap is for generating the x-axis values and the second-level bootstrap is for the estimation of the density of $T$ at these corresponding x-axis values. The right figure displays the estimated bootstrap likelihood curve.

Figure 1 is an ilustration of the bootstrap likelihood construction. In the next section, we will use the bootstrap likelihood to develop an algorithm to address Bayesian inference in the spirit of Mengersen, Pudlo and Robert (2013).

## 3. Bayesian computation via bootstrap likelihood

Our algorithm is described in algorithm 1 where $BL(\theta_i)$ denotes the estimation of the bootstrap likelihood in the point $\theta_i$

---

**Algorithm 1.** *Bayesian Computation via bootstrap likelihood*

---

*Estimate the bootstrap likelihood curves of parameters with the samples as described in the previous section.*
***for*** $i = 1$ *to* $M$ *do*

1. *Generate* $\theta_i$ *from the prior distribution* $\pi(\cdot)$
2. *Set the weight* $w_i = BL(\theta_i)$

***end for***

---

The output is a sample of size $M$ of parameters with associated weights, which operate as an importance sampling output. This means that a posterior sample of simulated parameters of size $N$ is sampled with replacement from the $M$ parameters with corresponding weights $w_i$'s. The bootstrap likelihood approach allows to define an algorithm with the same structure of the one defined in Mengersen, Pudlo and Robert (2013). As pointed out in the previous Section, the convergence rate of the bootstrap likelihood is the same of the empirical likelihood. Furthermore, in contrast with the empirical likelihood method, the bootstrap likelihood doesn't require any set of subjective constrains taking advantage from the bootstrap methodology. This makes the algorithm an automatic and reliable procedure where only a few parameters need to be specified. Another benefit of using the bootstrap likelihood instead of the empirical likelihood is that the construction of bootstrap likelihood could be done once and not at every iteration as the empirical likelihood. This leads to significant improvement in the computing time when different priors are compared.Throughout the paper, we will denote the algorithm of Mengersen, Pudlo and Robert (2013) as $BC_{el}$ and our algorithm as $\mathrm{BC}_{bl}$.

As a toy illustration of the method, we apply the $\mathrm{BC}_{bl}$ algorithm to a normal distribution with known variance (equal to 1). Clearly, the parameter of interest is the mean and we can see in Figure 2 the fitting of the posterior distribution. In this experiment, the computing time of $\mathrm{BC}_{bl}$ algorithm is much less than $BC_{el}$ method. The main reason is that the estimation of the parameter (sample mean in this case) is explicit and straightforward, without need of numerical estimation algorithms.

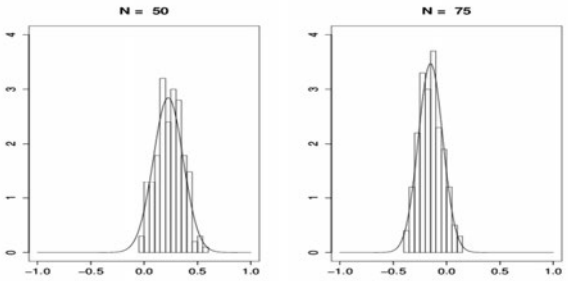


Figure 2: Comparison of the true posterior on the normal mean (solid lines) with the empirical distribution of weighted simulations resulting from $\text{BC}_{bl}$ algorithm. The normal sample sizes are 50 and 75 respectively, the number of simulated $\theta$'s is 200.

In the next Section, the performance of the bootstrap likelihood approach is explored in several examples. In particular, we will see how to manage the parameter estimation in the nested bootstrap. As we will see, this step of the methodology can vary with the problem at the hand.

## 4. Numerical Illustration

### *4.1. Dynamic Models*

As mentioned in Section 2, one way to deal with the dependence in dynamic models, is through the application of the bootstrap procedure to the unobserved i.i.d. residuals. For example, we test the $GARCH(1,1)$ model which has the following expression (see Bollerslev (1986)):

$$y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0,1), \quad \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

under the constraints $\alpha_0, \alpha_1, \beta_1 > 0$ and $\alpha_1 + \beta_1 < 1$. An exponential $Exp(1)$ and a Dirichlet $Dirich(1,1,1)$ priors are assumed, respectively, on $\alpha_0$ and $(\alpha_1, \beta_1, 1-\alpha_1-\beta_1)$. In order to compare with $BC_{el}$, we set the constrains for the empirical likelihood as Mengersen, Pudlo and Robert (2013). The numbers of first and second level of bootstrap replicates are respectively, $K = 100$ and $L = 1000$. For each bootstrap replicate of time series, R function `garch` is used for the estimation of parameters. Despite the lack of stability of this function, in Figure 3 we can see that the $\text{BC}_{bl}$ algorithm is performing better with respect to the $BC_{el}$ algorithm in term of parameter estimation, the ability to catch the correct range of $\alpha_0$, $\alpha_1$ and $\beta_1$. Moreover, we can observe that the $\text{BC}_{bl}$ algorithm produces estimates with less variability than the $BC_{el}$ algorithm. Finally, by the point of view of the computational time, we noted in our experiments that our approach is faster than the empirical likelihood one. This is not surprising because the bootstrap likelihood procedure depends heavily on the parameter

estimator. In this example, the R function **garch** provides a fast estimation of the model parameters and consequently a better computational time.

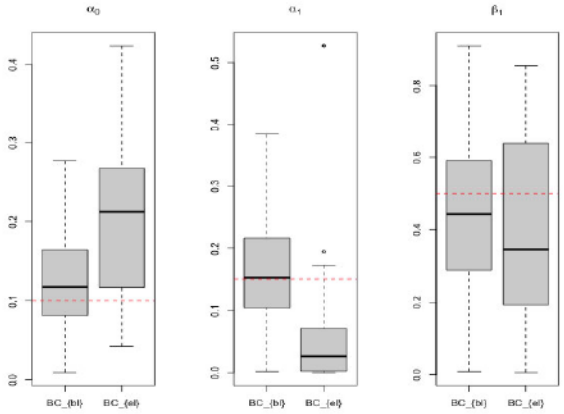

Figure 3: Comparison of evaluations of posterior expectations. (with true values in dashed lines) of the parameters $(\alpha_0, \alpha_1, \beta_1)$ of the $GARCH(1,1)$ model with 300 observations.

### *4.2. Population Genetics*

Mengersen, Pudlo and Robert (2013) applied the empirical likelihood procedure in a population genetics setting to model the evolutionary history of species. In this framework, it is important to be able to study the so called microsatellite distribution. Specifically, the the intra-locus likelihood is approximated by a product over all pair of genes in the sample at a given locus. The $k$-th locus likelihood is defined as

$$l_2\left(y^k|\phi\right) = \prod_{i<j} l_2\left(y_i^k, y_j^k|\phi\right)$$

where

$$l_2\left(y_i^k, y_j^k|\phi\right) = \begin{cases} \frac{\rho(\theta)^{|y_j^k - y_i^k|}}{\sqrt{1+2\theta}} & \text{if same deme} \\ \frac{e^{-\tau\theta}}{\sqrt{1+2\theta}} \sum_{k=-\infty}^{\infty} \rho(\theta)^{|k|} I_{|y_i^k - y_j^k|-k}(\tau\theta) & \text{if different deme} \end{cases}$$

We compare our proposal with $BC_{el}$ in a simple example described in the first experiment of Mengersen, Pudlo and Robert (2013), whose evolutionary scenario is displayed in Figure 4. Anyway, we refer to Mengersen, Pudlo and Robert (2013) for all the details of this model.

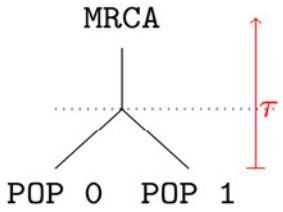


Figure 4: Evolutionary scenario of genetic experiment.

In this scenario, there are two parameters of interest $\tau$ and $\theta$. Specifically, $\tau$ is the time at which the two populations diverged in the past and $\theta/2$ is the mutation rate of the mutations at a given locus. The simulated datasets are made of ten diploid individuals per population genotyped at fifty independent loci. We first compare the marginal posterior distribution of the parameters $\theta$ and $\tau$ both with the $BC_{el}$ and $\text{BC}_{bl}$ sampler. Figure 5 suggests that $BC_{el}$ has difficulties eliminating the tails of both posterior distributions and $\text{BC}_{bl}$ is more accurate in terms of the shape.

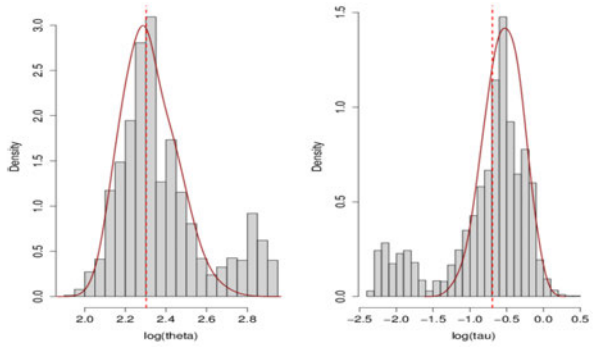


Figure 5: Comparison of the marginal distributions obtained by the $BC_{el}$ and the $\text{BC}_{bl}$ sampler. The histogram is sampled using $BC_{el}$ and the curve is the result of $\text{BC}_{bl}$.

Mengersen, Pudlo and Robert (2013) further suggest the incorporation of empirical likelihood in the adaptive multiple importance sampling (AMIS) to speed up the computation in this example, meanwhile to improve the performance of the basic $BC_{el}$ algorithm by adjusting the importance weights. The bootstrap likelihood could also be incorporated in the same way. However, Figure 6 shows that AMIS improves substantially the results computed with the basic $BC_{el}$ sampler, but not so much with respect to the $\text{BC}_{bl}$ sampler. Even in the case of $\tau$ estimation, integration with AMIS makes the posterior distribution calculated with $\text{BC}_{bl}$ deviate more from its true value. It appears that the basic $BC_{bl}$ sampler is enough capable of building a reasonable posterior. This suggests that is unnecessary to introduce the AMIS in the bootstrap likelihood setting.

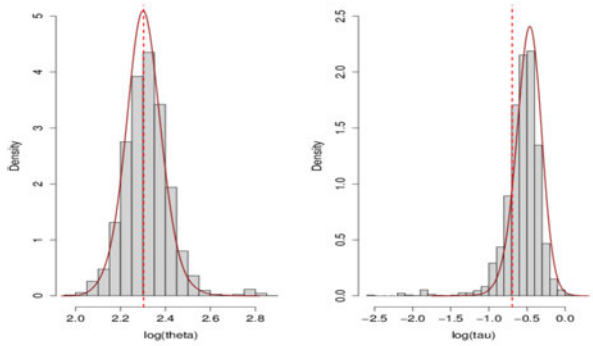


Figure 6: Comparison of the $BC_{el} - AMIS$ and the $BC_{bl} - AMIS$ sampler. The histogram is sampled using $BC_{el} - AMIS$ and the curve is the result of $BC_{bl} - AMIS$.

*4.3. Bivariate Beta distribution*

In this example, we consider the application of the $\text{BC}_{bl}$ to parameter estimation for the 5-parameter bivariate beta distribution proposed by Arnold and Ng (2011). This model is an extension of the 3 parameter specification of Olkin, Liu (2003), which only allows for positive correlation. The 5-parameter model is thus more flexible as it allows for both positive and negative correlations. However, the increase of the number of parameters from 3 to 5 pays the price of not having a closed form for the joint density function. To address the estimation problem, Arnold and Ng (2011) propose the so called modified maximum likelihood estimation (MMLE) approach, which combines MLE for the marginals with a method of moments estimate. We use the MMLE as the estimation approach in each bootstrap level of the bootstrap likelihood estimation.

The 5-parameter bivariate beta distribution $(X_1, X_2)$ can be described as follows:

$$X_1 = \frac{U_1 + U_3}{U_1 + U_3 + U_4 + U_5}$$
$$X_2 = \frac{U_2 + U_4}{U_2 + U_3 + U_4 + U_5}$$

where $U_i,\ i = 1, ..., 5$ is the random variable with gamma distribution $G(\alpha_i, 1)$ with shape parameter $\alpha_1$ and scale parameter 1. The marginal distributions of $X_1$ and $X_2$ are beta distributions with parameters $(\alpha_1 + \alpha_3, \alpha_4 + \alpha_5)$ and $(\alpha_2 + \alpha_4, \alpha_3 + \alpha_5)$, respectively.

Crackel and Flegal (2014) apply the ABC to address parameter inference for the bivariate beta distribution. We compare the performance of $BC_{el}$, $\text{BC}_{bl}$ and ABC using posterior mean and MSE in the following four parameter settings, $A_1 = (1, 1, 1, 1, 1)$, $A_2 = (1, 2, 2, 5, 0.1)$, $A_3 = (5, 2, 0.5, 0.5, 1)$ and $A_4 = (10, 10, 0.1, 0.1, 1)$, based on the fact that each parameter setting represents one kind of correlation relationship between $X_1$ and $X_2$. Several moments of the marginal beta distributions are known in Arnold and Ng (2011), so the empirical

likelihood is computed under the moment constrains $E(X_1)$, $E(X_1^2)$, $E(X_2)$, $E(X_2^2)$ and another expection which is used in MMLE, $E\left[\frac{(1-X_1)(1-X_2)}{X_1X_2}\right]$. For the ABC configuration, we choose the same sufficient summary statistics as Crackel and Flegal (2014), which are $S_1(\boldsymbol{x}) = \frac{1}{n}\sum lnx_{1i}$, $S_2(\boldsymbol{x}) = \frac{1}{n}\sum ln(1-x_{1i})$, $S_3(\boldsymbol{x}) = \frac{1}{n}\sum lnx_{2i}$, $S_4(\boldsymbol{x}) = \frac{1}{n}\sum ln(1-x_{2i})$ and $S_5$ the Pearson correlation between $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$. The distance function is chosen as $\rho(S(\tilde{\boldsymbol{x}}), S(\tilde{\boldsymbol{y}})) = \sum_{i=1}^{5} |S_i(\tilde{\boldsymbol{x}}) - S_i(\tilde{\boldsymbol{y}})|$, and the tolerance is set as 0.6 on the balance of accuracy and computation cost.

In order to do a proper comparison, we put a gamma prior $G(\tilde{\alpha_i}, 1)$ on parameter $\alpha_i$, where $\tilde{\alpha_i}$ is the real value of $\alpha_i$. The reason of selecting this "informative" prior is that in some parameter settings, especially in $A_4$ where the parameters has large discrepancy, the universal prior as in Crackel and Flegal (2014) for all the parameters leads to catastrophic estimation results for $BC_{el}$ and ABC, thus it is not so informative to do a comparison under that prior. We also observe that ABC and $BC_{el}$ are highly sensitive to the prior choice, while the effect of prior on $\text{BC}_{bl}$ are not as noticeable.

Table 1: Comparison of $\text{BC}_{bl}$, $BC_{el}$ and ABC

| $A_1 = (1,1,1,1,1),\ r = -0.29$ | | | | | | | $A_1 = (1,2,2,5,0.1),\ r = -0.75$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\text{BC}_{bl}$ | | $BC_{el}$ | | ABC | | $\text{BC}_{bl}$ | | $BC_{el}$ | | ABC | |
| Parameter | Mean | MSE | Mean | MSE | Mean | MSE | Mean | MSE | Mean | MSE | Mean | MSE |
| $\hat{\alpha_1}$ | 0.886 | 0.144 | 1.012 | 0.160 | 1.059 | 0.504 | 1.022 | 0.134 | 0.947 | 0.125 | 0.914 | 0.362 |
| $\hat{\alpha_2}$ | 0.934 | 0.158 | 0.967 | 0.211 | 1.101 | 0.637 | 1.611 | 0.651 | 1.681 | 0.684 | 2.141 | 1.826 |
| $\hat{\alpha_3}$ | 1.123 | 0.064 | 1.000 | 0.073 | 1.215 | 0.311 | 1.926 | 0.067 | 1.948 | 0.066 | 2.081 | 0.343 |
| $\hat{\alpha_4}$ | 1.193 | 0.133 | 1.158 | 0.135 | 1.321 | 0.419 | 5.190 | 0.511 | 5.085 | 0.493 | 5.236 | 3.030 |
| $\hat{\alpha_5}$ | 0.908 | 0.069 | 0.907 | 0.092 | 1.079 | 0.159 | 0.028 | 0.007 | 0.033 | 0.007 | 0.040 | 0.014 |
| $A_1 = (5,2,0.5,0.5,1),\ r = 0.28$ | | | | | | | $A_1 = (10,10,0.1,0.1,1),\ r = 0.71$ | | | | | |
| | $\text{BC}_{bl}$ | | $BC_{el}$ | | ABC | | $\text{BC}_{bl}$ | | $BC_{el}$ | | ABC | |
| Parameter | Mean | MSE | Mean | MSE | Mean | MSE | Mean | MSE | Mean | MSE | Mean | MSE |
| $\hat{\alpha_1}$ | 5.023 | 0.764 | 4.916 | 1.831 | 5.340 | 3.840 | 9.440 | 2.278 | 8.914 | 5.158 | 10.075 | 5.995 |
| $\hat{\alpha_2}$ | 2.169 | 0.208 | 2.162 | 0.340 | 2.372 | 1.372 | 9.183 | 2.088 | 8.858 | 3.701 | 9.900 | 6.227 |
| $\hat{\alpha_3}$ | 0.434 | 0.081 | 0.555 | 0.105 | 0.475 | 0.187 | 0.073 | 0.011 | 0.175 | 0.059 | 0.050 | 0.015 |
| $\hat{\alpha_4}$ | 0.553 | 0.031 | 0.496 | 0.062 | 0.462 | 0.114 | 0.082 | 0.012 | 0.153 | 0.035 | 0.045 | 0.016 |
| $\hat{\alpha_5}$ | 1.081 | 0.041 | 1.073 | 0.077 | 1.246 | 0.163 | 0.980 | 0.030 | 0.998 | 0.036 | 1.213 | 0.094 |

We can see from table 1 that three algorithms perform more or less the same in terms of being able to capture the true parameter value. But the $\text{BC}_{bl}$ has the least MSE in most cases and ABC tends to give the largest MSE. For example, Figure 7 displays the boxplots of all the 5 parameters evaluated with $ABC$, $\text{BC}_{bl}$ and $BC_{el}$ respectively under the parameter

setting $A_3$.

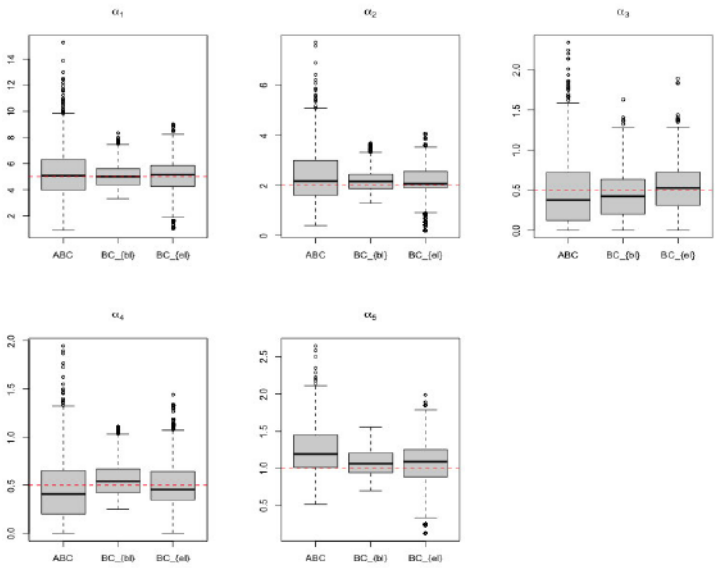


Figure 7: Comparison of evaluations of posterior expectations (with true values in dashed lines) of the parameters $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ of the bivariate distribution with 100 observations.

### *4.4. Ising and Potts Model*

Ising and Potts models are discrete Gibbs random field models originating in statistical physics, which are now widely used in statistics for applications in spatial modeling, image processing, computational biology, and computational neuroscience. Consider the simple case of a random field where the pixels of the image $\boldsymbol{x}$ can only take two colors (white and black, say). Let $\{\boldsymbol{x} = x_{ij} : (i,j) \in D\}$ denote the observed binary data, where $x_{ij}$ is a spin and $D$ is an $M \times N$ lattice indexing the spins. The conditional distribution of a pixel is then Bernoulli, with the parameter being a function of the number of neighboring pixels that have the same value. It is defined as

$$f(x_{ij} = k | x_{n(i,j)}) \propto \exp(\beta n_{i,j}^k), \qquad \beta > 0, \quad k = 0, 1$$

where

$$n_{i,j}^k = \sum_{l \in n(i,j)} \mathbb{I}_{x_l = k}$$

is the number of neighbours of $x_{ij}$ with color $k$ and $n(i,j) = \{(i+1,j),(i-1,j),(i,j+1),(i,j-1)\}$ is the defined neighborhood structure. In Statistical Mechanics, $\beta$ is a strictly positive parameter which can be interpreted as the inverse of the temperature. The Ising

model is defined through these full conditionals

$$f(x_{ij} = 1|x_{n(i,j)}) = \frac{\exp(\beta n^1_{i,j})}{\exp(\beta n^0_{i,j}) + \exp(\beta n^1_{i,j})}$$

and the joint distribution therefore satisfies

$$f(\boldsymbol{x}) \propto \exp(\beta \sum_{(i,j)\sim(i',j')} \mathbb{I}_{\{x_{ij}=x_{i'j'}\}})$$

where the summation is taken over all the neighbour pairs. The Potts model is the natural extension of the Ising Model where more than two colors are considered, see Marin, Robert (2014).

The normalizing constant $Z(\beta)$ of the above distribution depends on $\beta$ and it is only available for very small lattices $D$, which becomes a major obstacle when inferring on $\beta$. The maximum pseudo-likelihood estimator (MPLE) Besag, J. (1977) provides a way to handle the problem. MPLE takes the value that maximizes the pseudo-likelihood function

$$L(\beta|\boldsymbol{x}) = \prod_{i=1}^{M}\prod_{j=1}^{N} f(x_{ij} = 1|x_{n(i,j)}, \beta)$$

We will adopt MPLE as the estimation tool to construct the bootstrap likelihood for $\beta$ later. Marin, Robert (2014) introduce ABC as a way to simulate the posterior distribution $\beta$. However, simulating a data set is unfortunately non-trivial for Markov random fields, as it usually requires a certain number of steps of an MCMC sampler.

We compare the performance of ABC and $\text{BC}_{bl}$ in a simulation dataset of size $25 \times 25$ where the true parameter $\beta$ is set as 0.5. The simulation is done using the Gibbs sampler, starting with a random configuration with each spin being drawn independently from $\{0, 1\}$, and then iterating for 200 Gibbs cycles. The sufficient statistic $S$ is

$$S(\boldsymbol{x}) = \sum_{(i,j)\sim(i',j')} \mathbb{I}_{\{x_{ij}=x_{i'j'}\}}$$

For assessing bootstrap likelihood, the moving block bootstrap technique and MPLE are employed. To preserve the spatial dependency of the data, the window length of the moving block is set as 5. The numbers of bootstrap replicates for the 1st level and 2nd level bootstrap are 100 and 200, respectively.

Figure 8 shows that the estimation carried with $\text{BC}_{bl}$ and ABC algorithms provides similar results. It is worth to mention that the $\text{BC}_{bl}$ has a computational cost which is less than ABC since the Gibbs sampling for the Ising model has a cost which increases quadratically

as the lattice grows. The same problem arises with Potts model where more than two colours are considered.

We conclude this section with a real data example. In particular, we consider a set of soil phosphate measurements collected during the Laconia Archaeological Survey in Greece (year 1987). A complete description of data can be found i.e. in Buck et al. (1988). This dataset has been analysed by using different techniques, for instance Buck et al. (1988) carried a Bayesian change-point analysis to describe the dataset. On the other hand, Besag et al. (1991) adopted a Bayesian image analysis approach. Recently, McGrory et al. (2009) studied the dataset with variational Bayes methods.

In this application, we use the moving block bootstrap and MPLE techniques in a similar way as the simulation study. The window length of the moving block is set as 8. The numbers of bootstrap replicates for the 1st level and 2nd level bootstrap are 100 and 100, respectively. The distribution of values of $\beta$ is shown in figure 9. It can be noticed that results are quite similar to those obtained in the recent paper of McGrory et al. (2009) who use a variational Bayes method.

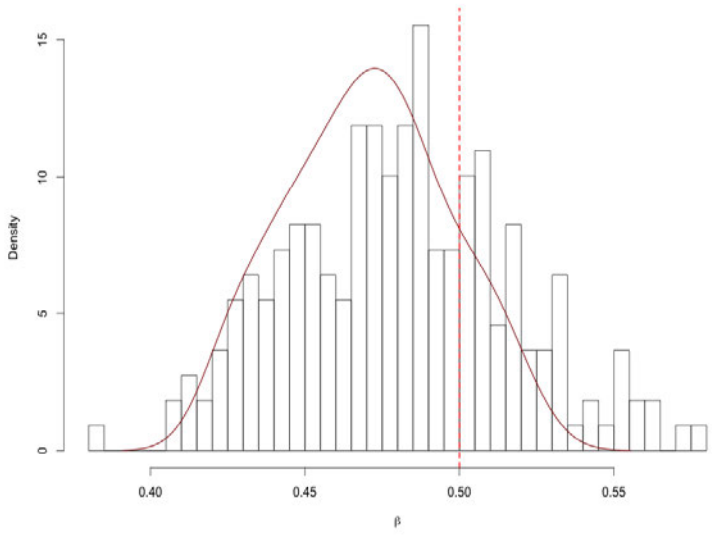


Figure 8: Comparison of the $BC_{bl}$ (curve) with the histogram of the simulations from ABC algorithm with $10^4$ iterations and a 1% quantile on the difference between the sufficient statistics as its tolerance bound $\epsilon$, based on the uniform prior $U(0, 2)$.

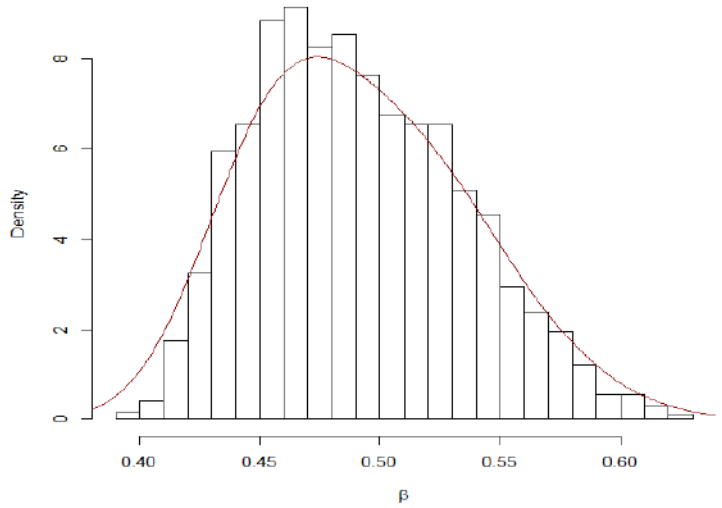


Figure 9: Histogram and density estimation of parameter $\beta$ after applying $BC_{bl}$ in the Laconia Archaeological data.

### *4.5. Conclusion*

In this paper, we introduced a bootstrap likelihood approach to address inference in a Bayesian setting. The sampling scheme has a structure which is similar to the recent work of Mengersen, Pudlo and Robert (2013). These type of algorithms can be used as an alternative to the standard ABC methods when difficulties arise in setting the parameters (distance, summary statistics and tolerance level). In particular, the empirical likelihood approach of Mengersen, Pudlo and Robert (2013) allows to avoid this problem but, on the other hand, requires to choose a set of constrains. Different choices could sensibly affect the inference and, consequently, the parameter estimation. The main advantage of the bootstrap likelihood approach is that it is an automatic procedure that does not require the careful choice of the constrains. In the paper, merits and problems of the new approach are discussed through simulation experiments. In particular, the method is tested on a dynamic model, population genetics and a bivariate beta distribution. Furthermore, in a random field context, an application to real data is provided.


### Acknowledgements

J. Miguel Marin's research has been supported by grant MTM2010-17323. Fabrizio Leisen's research has been supported by the Royal Society Grant "Bayesian Computational methods with Bootstrap Techniques" and by the European Community's Seventh Framework Programme[FP7/2007-2013] under grant agreement no:630677. The authors are very grateful to Pierre Pudlo for all the suggestions about the population genetic example.

## Appendix

*Composite likelihood*

One of the alternatives of MLE inside the nested bootstrap estimation is to use the Maximum Composite Likelihood Estimator (MCLE). Composite likelihood is an inference function derived by the product of a collection of suitable low dimensional component likelihoods, for instance, conditional or marginal densities. The development of composite likelihood is generally motivated by the issue of computational feasibility arising in the application of the likelihood-based method in high-dimensional data analysis and models with complicated structure. Composite likelihood inherits many of the good properties of inference based on the full likelihood function, yet is more easily implemented with high-dimensional data sets. There is a wide range of application areas, such as geostatistics, space-time models, time series, population genetics and so on.

Let $y = (y_1, ..., y_n)$ be a random sample from $Y_i \sim f(y_i; \theta)$, where $y_i \in \mathcal{Y} \subseteq \mathbb{R}^q$ and let $\{A_1(y_1), ..., A_2(y_2)\}$ be a set of marginal or conditional events with associated likelihoods $\mathcal{L}_k(\theta; y_i) \propto f(y_i \in A_k; \theta)$, then a composite likelihood is defined as

$$\mathcal{L}_C(\theta; y) = \prod_{i=1}^{n} \prod_{k=1}^{K} \mathcal{L}_k(\theta; y_i)^{\omega_k}$$

where $\omega_k$ are non-negative weights. When the events $A_k(y_i)$ are defined in terms of pairs of bivariate marginal densities $f_{st}(y_{is}, y_{it}; \theta)$, the so-called pairwise likelihood is given by

$$\mathcal{L}_P(\theta; y) = \prod_{i=1}^{n} \prod_{s,t=1, s \neq t}^{q} f_{st}(y_{is}, y_{it}; \theta)^{\omega_{st}}$$

.

## References


Barry, C.A. and Ng, H.K.T. (2011). Flexible bivariate beta distributions. J. Mult. Anal. 102, 1194-1202.

Beaumont, M., Zhang, W. and Balding, D. (2002). Approximate Bayesian computation in population genetics. Genetics 162(4), 2025-2035.

Besag, J. (1977). Efficiency of pseudolikelihood estimation for simple Gaussian fields. Biometrika, 616-618.

Besag, J., York, J. and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). Ann. Inst. Stat. Math. 43, 15.

Bollerslev, T. (1986). Generalized Autorregressive Conditional Heteroskedasticity. J. Econometrics 31, 307-327.

Buck, C.E., Cavanagh, W.G. and Litton, C.D. The spatial analysis of site phosphate data. In: Rhatz, S.P.Q. (ed.) Computer Applications and Quantitive Methods in Archeology. British Archaeological Reports, International Series, vol. 446. BAR, Oxford, 1988.

Crackel, R.C. and Flegal, J.M. (2014). Approximate Bayesian computation for a flexible class of bivariate beta distributions. `http://arxiv.org/abs/1402.1782`.

Davison, A.C., Hinkley, D.V. and Worton, B.J. (1992). Bootstrap likelihoods. Biometrika 79(1), 113-130.

Davison, A.C. and Hinkley, D.V. Bootstrap methods and their application. Cambridge university press, 1997.

Dean, T., Singh, S., Jasra, A. and Peters, G. (2014). Estimation of HMMs with Intractable Likelihoods. Scand. J. Statist. (to appear).

Drovandi, C. and Pettitt, A. (2010). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. Biometrics 67(1), 225-233.

Fearnhead, P. and Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. J. Royal Statistical Society (B) 74(3), 419-474.

Marin, J.M. and Robert, C.P. Bayesian Essentials with R. Springer, 2014.

McGrory, C.A., Titterington, D.M., Reeves, R. and Pettitt, A.N. (2009). Variational Bayes for estimating the parameters of a hidden Potts model. Stat. Comput. 19, 329-340.

McKinley, T., Cook, A. and Deardon, R.(2009). Inference in epidemic models without likelihoods. Int. J. Biostat. 5(1), 24.

Mengersen, K.L., Pudlo, P. and Robert, C.P. (2013). Bayesian computation via empirical likelihood. Proceedings of the National Academy of Sciences 110(4), 1321-1326.

Olkin, I. and Liu, R. (2003). A bivariate beta distribution. Stat. Prob. Letters 62(4), 407-412.

Del Moral, P., Doucet, A. and Jasra, A. (2012). An adaptive sequential Monte Carlo method for approximate Bayesian computation. Stat. Comput. 22, 1009-1020.

Sisson, S.A., Fan, Y. and Tanaka, M. (2007). Sequential Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. 104(6), 1760-1765.